

ICS 35.240
CCS L 70



**NATIONAL STANDARD OF THE
PEOPLE'S REPUBLIC OF CHINA**

GB/T 47517-2026

**Information technology - Technical specification for sign
language digital humans**

信息技术 手语数字人技术规范

Issued on: April 30, 2026

Implemented on: November 1, 2026

**Issued by: State Administration for Market Regulation;
Standardization Administration of the PRC**

Table of Contents

6 Technical Requirements

6.1 Input Module

6.1.3 Voice Input

6.1.4 Image Input

6.2 Preprocessing Module

6.2.1 Speech Recognition

6.3 Sign Language-Driven Expression

6.4 Character Generation

7 Test Methods

7.2 Input Module

7.2.3 Voice Input

7.3 Preprocessing Module

7.3.1 Speech Recognition

7.4 Character Portrayal

7.5 Sign Language-Driven Expression

8 Unicode Transformation Format-8-bit (8-bit variable length) WAV. Waveform Audio

Foreword

GB/T 47517-2026 | Information technology - Technical specifications for sign language digital human

GB/T 47517-2026 English version. Information technology - Technical specifications for sign language digital human National Standards of the People's Republic of China ICS 35.240 CCS L

70 Information Technology Sign Language Digital Human Technical Specifications
Released on April 30, 2026 Implemented on 2026-11-

01 State Administration for Market Regulation The State Administration for Standardization issued a statement.

1.Scope This document specifies the technical framework and requirements for sign language digital humans, and describes the corresponding testing methods. This document applies to the development and testing of sign language digital humans.

4.Abbreviations The following abbreviations apply to this document. AAC. Advanced Audio

Coding ASR. Automatic Speech Recognition BMP. Bitmap FLAC. Free Lossless Audio Codec JPEG. Joint Photographic Experts Group MP3.MPEG Audio Layer III OCR. Optical Character Recognition PNG. Portable Network Graphics UAC. USB Audio Class UTF-

6.1 Input Module

6.1.1 Supported Forms A sign language digital human should at least support text input, and preferably also support voice input via uploaded audio files, voice input for natural dialogue, and image input. The sign language digital human should support interaction in Simplified Chinese and Mandarin, and can support other languages and scripts on top of that.

Note. Unless otherwise specified, the technical requirements and test methods in this document are in Simplified Chinese and Mandarin.

6.1.2 Text Input The requirements are as follows:

- a) It should at least support UTF-8 encoding format;
- b) The maximum input length of supported text should be no less than 200 characters.

6.1.3 Voice Input

6.1.3.1 Voice Input for Uploaded Audio The requirements are as follows:

- a) It should support at least the mainstream audio formats such as WAV, MP3, AAC, and FLAC;
- b) It should support audio sampling rates in the range of at least 8 kHz to 48 kHz;
- c) It should support at least mono audio;
- d) The maximum supported audio input duration should be no less than 120 seconds.

6.1.3.2 Voice Input for Natural Dialogue The requirements are as follows:

- a) Audio capture via UAC standard microphones (built-in or external) should be supported.
- b) At least one of the following voice activation mechanisms should be supported. --Starts after the user clicks the "Start" button; --Silence detection activated; --Hot words awaken.
- c) At least 120 seconds of continuous voice input should be supported.

6.1.4 Image Input

- a) It should support at least the mainstream image formats such as BMP, JPEG, and PNG;
- b) It should support image resolutions of at least 800 px x 600 px;
- c) The maximum supported single file size should be no less than 5 MB.

6.2.1 Speech Recognition

6.2.1.1 Speech Recognition of Uploaded Audio If the sign language digitizer supports uploading audio input, the log should record the start time of the sign language digitizer's speech recognition and the completion time of the recognition. The time point of formation and the identification results. The speech recognition performance requirements for uploaded audio are as follows:

- a) The real-time factor should be less than or equal to 1.0;
- b) The accuracy rate should be no less than 85%.

6.2.1.2 Speech Recognition for Natural Dialogue If the sign language digital human supports natural dialogue input, then the time when the first audio frame arrives at the sign language digital human and the time when the audio frame ends should be recorded in the log. The time point when the frame arrives at the sign language digital human, the time point when speech recognition ends, and the recognition result. The speech recognition performance requirements for natural dialogue are as follows:

- a) The real-time performance factor should not exceed 1.2;
- b) The accuracy rate should be no less than 85%.

6.2.2 Optical Character Recognition If the sign language digital human supports image input with text, the time it takes for the image data to fully arrive at the sign language digital human should be recorded in the log. The time point of output for dot and optical character recognition results, and the recognition results. The performance requirements for optical character recognition are as follows:

- a) The response efficiency ratio should not exceed 0.8;
- b) The accuracy rate should be no less than 85%.

6.3 Sign Language-Driven Expression

6.3.1 Standardization of Sign Language Movements Each sign language gesture should be approximated to the level of accuracy required by Chapter 7 of GF 0020-2018. Fraser distance and normalization should be used. The dynamic time-normalized distance measurement, after optimization, is used to assess the standardization of sign language movements, with the following requirements.

- a) The average Fraser distance for each sign language gesture, body posture, and facial expression should be less than 20;
- b) The normalized dynamic time warp distance for each action should be less than 0.5.

6.3.2 Sign Language Comprehension The degree to which written translations into sign language can be understood by deaf people. Subjective evaluation was used, with the

assessment team's average comprehension score and more than half of the test results being considered. The evaluators' comprehension score should be no less than 70 points.

6.3.3 Real-time performance When the input is less than or equal to 200 characters, the sign language digital human should start up with a delay of less than 3 seconds.

6.3.4 Speed Control It supports different sign language expression speeds, with at least three speed levels: fast, medium, and slow.

6.3.5 Video Stream Performance The sign language video stream should support at least 25 frames per second. In a 1280px x 720px scene, the network bandwidth requirement should not exceed

3 Mbit/s. In a 1920px x 1080px scenario, the network bandwidth requirement is no higher than

6.4 Character Generation

6.4.1 Face The facial requirements for sign language digital humans are as follows:

- a) Facial contours should be clear and sharp, without jagged edges, blurriness, distortion, or missing parts;
- b) It should have a standard facial structure, with harmonious proportions, and avoid facial distortion;
- c) Basic facial expressions and mouth shape changes should be supported to complement sign language semantic expression;
- d) There should be no missing, damaged, or obvious seams in the facial surface texture;
- e) Facial expressions should be natural and smooth, avoiding pauses, misalignments, or stiffness.

6.4.2 Hands The hand gestures required for a sign language digital human are as follows:

- a) The hand model should fully represent the structure of the five fingers, including the palm, the web between the thumb and index finger, and the five fingers. Each finger should have three joint segments to ensure... The structure is complete;
- b) The hand model should support natural and accurate finger movements, including bending, extension, adduction, abduction, and rotation required by sign language. Common movements, clear trajectory, coordinated hand movements, accurate movements and positions, and no bone misalignment or twisting during the movement. Curves, clipping, or sudden changes in movement;
- c) The surface of the hand should have sufficient visual detail, including nail shape, skin texture, fingertip creases, palm lines, etc.